testified that, for order completion interval, BellSouth can report compliant support even though it is providing discriminatory support in reality. The retail analog for Order Completion Interval-UNE Loops is Retail Residence and Business Dispatch. According to witness Bursh, a significant percent of the UNE Loop observations could be UNE analog loops, which are all dispatch-in. Dispatch-in signifies that the work is done within the Central Office. Dispatch usually refers to service where the work is done in the field or outside of the central office. Witness Bursh states that "work done within the central office has a shorter interval than work done away from the central office. Given that the retail analog [for Order Completion Interval-UNE Loops] is designed as Retail Residence and Business Dispatch, BellSouth would always be providing longer interval for itself (compliant support) for this example primarily because the retail analog is inappropriate."

In its brief, the ALEC Coalition states:

> As to benchmarks, the dispute between the parties is more clearly drawn. Again, BellSouth has chosen benchmark values that it believes are appropriate based upon the Louisiana and Georgia proceedings, and which are the same as those approved by the Georgia Commission. In contrast, the ALECs have proposed benchmarks that range from 95 to 100 percent (i.e., perfection). The ALECs have proposed no benchmark below 95 percent. In making their proposal, the ALECs have obviously deviated from what was accepted in Louisiana and Georgia. The specific values of the benchmarks proposed by Ms. Kinard on behalf of the ALECs are not substantively supported anywhere in her testimony. Further, Ms. Kinard admitted upon cross-examination that the ALECs have no analysis or study to support the conclusion that a 95 percent benchmark is the minimum 'that would allow ALEC a meaningful opportunity to compete.'"

At the hearing when witness Coon was asked how BellSouth determined what the appropriate benchmarks should be, his response was that most of the benchmarks proposed here are those that have been ordered in the Georgia Commission. He testified that, while BellSouth may not believe that a benchmark is

appropriate, it is what was ordered in another jurisdiction. He could not provide any factual basis for establishment of the BellSouth-proposed benchmarks.

BellSouth witness Coon argues that Witness Kinard's comments suggest that this Commission should adopt the ALEC plan not the BellSouth plan. Witness Coon notes that witness Kinard simply presents her analogs and benchmarks without any critical analysis to support the conclusions she has reached. Witness Coon notes that its recommendations regarding benchmarks and analogs are a result of several years of work and have been conformed to the results reached in Georgia. While BellSouth agrees with the principle that simply having another state approve something does not necessarily mean it is appropriate for Florida, the fact that Georgia has approved these analogs and benchmarks should bear some weight.

DECISION

We agree with BellSouth that many years of work have gone into the development of the benchmarks and analogs proposed by BellSouth.

The ALEC Coalition testimony specifies one example of how, because of disaggregation, the analogs proposed by BellSouth are inappropriate. Witness Bursh identifies that the BellSouth proposal for Average/Order Completion Interval-USE Loops analog is Retail Residence and Business Dispatch. Witness Bursh proposes that many of the USE Loops in this category may be analog loops, which are not dispatched outside the central office. Witness Bursh believes it would be inappropriate to compare the aggregate category of UNE Loop to Retail Residence and Business Dispatch since BellSouth would conceivably be providing longer intervals for itself.

We agree that this level of aggregation is inappropriate and have made changes to the aggregation as specified above. As a result of creating more levels of disaggregation for compliance purposes, the analogs will also be more disaggregated. The appropriate benchmarks and analogs are shown in relation to the disaggregation specified above, in Attachment 7.

Using the example provided by the ALEC Coalition for Average/Order Completion Interval, there is no aggregate UNE Loops category in this Order. Loops would be segregated by analog and digital and by design and non design. Specifically, a two-wire analog Loop-Design would be compared to retail residence and business dispatch, while a nondesign two-wire analog loop would be compared retail residence and business (POTS excluding switch based orders) for compliance purposes. We find that these analogs are appropriate.

As to benchmarks, we agree with the ALEC Coalition that benchmarks set below 90 or 95 percent do not generally allow the ALECs a meaningful opportunity to compete. We are increasing many of the benchmarks that are set below this level for both reporting and compliance purposes.

XVII. ROOT CAUSE ANALYSIS

Herein, we consider whether BellSouth should be required to perform a root cause analysis. ALECs contend that if a failure occurs twice in three consecutive months, a root cause analysis is necessary to identify problems. BellSouth argues that it is an expensive, time-consuming process that is not always necessary.

Arguments

BellSouth witness Coon defines "root cause analysis" as an often formalized, comprehensive, and detailed investigation of all the component activities related to the delivery of the service in question. A root cause analysis may include participation by all BellSouth entities involved in the delivery of the service and include not only problem identification but also the development and implementation of solutions.

Witness Coon believes that BellSouth should never be required to perform a root cause analysis. He believes that BellSouth has the information necessary to identify problems and the incentive, by virtue of enforcement penalties, to correct those problems. He does not believe BellSouth, nor this Commission, should be required to use valuable resources on issues already addressed under a self-effectuating remedy plan.

As explained by the witness, a root cause analysis is an investigation of all component activities related to the delivery of a service to an ALEC identified as being inferior. BellSouth argues that the Performance Assessment Plan adopted by this Commission should not impose a requirement that BellSouth conduct a root cause analysis of a continuing source of disparity. Witness Coon states that the ALECs have failed to demonstrate such a need.

ALEC witness Bursh states that "a root cause analysis is a useful procedure for building action plans to remedy unacceptable performance and should be incorporated within a performance measurement system . . ." She also states that procedures, such as root cause analyses, which could potentially remedy recurrence of failures, are definitely essential.

Witness Bursh further states, "[t]he Georgia Public Service Commission Order stated that BellSouth must perform a "root cause analysis" and file with the Commission a corrective action plan within 30 days of the failure. The root cause analysis would be triggered if any measure fails twice in any three consecutive months in a calendar year."

DECISION

Witness Coon does not believe that BellSouth should be required to perform a root cause analysis if a self-effectuating enforcement plan is in place. We agree that conducting root cause analyses could become burdensome, using valuable resources of BellSouth and this Commission.

The ALEC Coalition comments in its brief that:

> "[i]t is ironic that BellSouth, who accused the ALEC Coalition of being interested primarily in constructing a plan that would become a revenue producing device, argues against a provision that would identify the source of the disparity, require that it be rectified, and in the process turn off the penalty payments.
>
> In a sense, there are some similarities between BellSouth's position and that of the ALEC Coalition. Like BellSouth, the ALEC Coalition believes it is

> imperative that the self-effectuating nature of the PAP not be disrupted. Specifically, the ALEC Coalition believes the conducting of a root cause analysis should not interfere with the timely payments called for by a BellSouth failure.

Witnesses Bursh and Ford believe that it is necessary to implement a root cause analysis whenever there are repeated failures. Witness Ford believes BellSouth should not perform this analysis unless it is required under the performance assessment plan.

We are concerned that requiring a root cause analysis at this time could hinder initial implementation of the Florida Performance Assessment Plan. We find the implementation of a self-executing enforcement program is incentive enough for BellSouth to perform an analysis if and when penalties are paid out.

XVIII. STATISTICAL METHODOLOGY

With the exception of the appropriate level of aggregation for purposes of determining compliant performance, the parties appear to agree in broad conceptual terms on the appropriate methodology. For measures with a retail analog, all of the proposed plans employ a statistical approach to assess compliance. Further, the parties believe that a special provision should be made for small sample sizes. The standard for measures which do not have a retail analog is a benchmark, and the parties advocate a "bright-line" or "stare and compare" approach to determine compliance, with an allowance for small sample sizes. As will be discussed later, the parties disagree on the appropriate benchmark table for small sample sizes.

Therefore, where the standard for a measure is a retail analog, we find that compliance shall be evaluated through a statistical process. Where the standard for a measure is a benchmark, we find that compliance shall be determined by a "bright-line" comparison, with an adjustment for small sample sizes.

A. Parity

There is much similarity among the parties' testimony regarding the appropriate definition of parity. According to BellSouth's witness Coon, the following definitions of parity by the FCC should apply:

> 1) where a retail analog exists, the BOC must provide access to a competing carrier in substantially the same time and manner as it provides to itself; 2) for those functions that have no retail analogue, the BOC must provide access that would offer an efficient carrier a meaningful opportunity to compete.

ALEC Coalition witness Bursh also states that "benchmarks are set at a level that provides ALECs with a meaningful opportunity to compete."

Z-Tel witness Ford believes that parity service, non-discriminatory service, and the same level of service are all synonymous. In addition, witness Ford believes that service needs to be non-discriminatory for all sizes of ALECs.

From a statistical standpoint, BellSouth witness Mulrow and Z-Tel witness Ford provide similar definitions of parity. Witness Mulrow states that the "null hypothesis is really that the means are equal and the standard deviations are equal." Witness Ford opines that the null hypothesis is a "zero-means difference." Due to the agreement among the parties that there is a need to balance Type I and Type II errors, there must be some deviation in practice from the theoretical null hypothesis. Nonetheless, we find that the null hypothesis shall be defined as closely as possible to this ideal, while still incorporating error probability balancing as all parties support.

Therefore, we find that BellSouth witness Coon's definition of parity shall be adopted. Where a measure has a retail analog, BellSouth shall provide access to a competing carrier in substantially the same time and manner as it provides to itself. For those functions that have no retail analog, BellSouth shall provide access that would offer an efficient carrier a meaningful opportunity to compete.

In discussing the appropriate statistical methodology, the parties have offered testimony which describes how the methodology may need to vary depending on whether the measure is a mean measure, a proportion measure, or a rate measure. In addition, there is the small sample size problem, and the issue of the appropriate level of aggregation for purposes of assessing compliance, which directly affects the selection of the appropriate statistical methodology.

BellSouth witness Mulrow explains how mean measures, proportion measures, and rate measures are different types of statistics. In a comparison of means, witness Mulrow testifies that the average of the BellSouth transactions in a "cell" is compared to the average or mean of the ALEC transactions. Some measures, however, are not expressed in terms of means. Witness Mulrow cites missed appointments as an example of a proportion measure, where the statistic is expressed as a percentage. He cites a rate measure (e.g., customer trouble report rate) as another example of a statistic which is not stated in terms of a mean. While proportion measures cannot exceed 1, a rate measure may exceed 1. For mean measures, witness Mulrow observes that the statistical approach must consider the BellSouth and ALEC means and the standard deviation of BellSouth's mean. In the case of proportion and rate measures, the proportion or rate is the only parameter to consider. Witness Mulrow states that "BellSouth cannot separately control the proportion [or rate] value and the variability about that value." According to witness Mulrow, ALEC Coalition witness Bell inappropriately uses the same statistical approach for mean, proportion, and rate measures in his direct testimony.

Before discussing how the statistical approach may need to vary to fit the nature of the measure (mean, proportion, or rate), the error probability balancing concept needs to be introduced. As will be discussed below, Type I and Type II errors are common parlance among statisticians, and there is agreement among the parties as to what constitutes Type I and Type II errors. Further, there is agreement among the parties on the need to balance these two types of errors in the context of a Performance Assessment Plan.

With Type I error, Commission staff witness Stallcup indicates that the statistical test shows that BellSouth is

providing non-compliant service when in fact it is providing compliant service. Similarly, ALEC Coalition witness Bell states that a Type I error occurs if the statistical test shows that "BellSouth is favoring its retail operations when, in fact, parity service exists." Finally, Z-Tel witness Ford states that Type I error occurs when there is a false conclusion that service is discriminatory. We find that all of these descriptions are conceptually identical.

With Type II error, witness Stallcup indicates that the statistical test shows that BellSouth is providing compliant service when in fact it is providing non-compliant service. According to ALEC Coalition witness Bell, "a Type II error occurs if the statistical test fails to indicate that BellSouth is favoring its retail operations when, in fact, a certain degree of disparity does exist." Z-Tel witness Ford describes Type II error as "fail[ing] to detect discrimination that actually exists." Once again, we find that all of these descriptions are conceptually the same.

Witness Stallcup describes the Balancing Critical Value technique as a means to equalize Type I and Type II errors such that the enforcement mechanism will not be biased towards BellSouth or the ALECs. He goes on to state that this approach has the "intuitive appeal of balancing the interests of both BellSouth and the ALECs." Z-Tel witness Ford offers similar testimony, expressed in terms of penalty payments:

> With Type I error, the ILEC pays penalties for false positives. With Type II error, the ILEC does not pay penalties when it does in fact discriminate. Both problems need to be addressed within the context of a performance plan.

BellSouth witness Taylor also speaks to the motivation for balancing, namely the "desire to hold the risk of Type I error (which would favor the ALEC at BellSouth's expense) at exactly the same level as the risk of Type II error (which would favor BellSouth at the ALEC's expense)." Thus, we find that there is much agreement on the balancing concept, albeit dispute over the appropriate value for the parameter delta which is required to implement the concept.

Witness Stallcup observes that the choice of statistical methodology is a function of the level of disaggregation. If BellSouth's method of disaggregating the enforcement measures is deemed appropriate, BellSouth's test statistic (Truncated Z) would be appropriate. Similarly, if this Commission adopts the ALEC's method of disaggregating the enforcement measures, the ALEC's test statistic (Modified Z) would be appropriate.

In addition, witness Stallcup explains that both tests, Modified Z and Truncated Z, begin in the same way with a Modified Z test being performed at the "cell" level. Under the Truncated Z, the cell level results are in turn aggregated. The truncation involves setting cell level Z scores to zero, if the ALEC received superior service. For a mean measure, a Z score is calculated by dividing the difference between the ALEC and ILEC means by the standard deviation of this difference. Based on the assumption that both samples were drawn from the same population, the Z score has a sampling distribution that approximates a Standard Normal (i.e., the bell-shaped probability distribution).

ALEC Coalition witness Bell and Z-Tel witness Ford agree that the Truncated Z is appropriate to aggregate homogeneous cells. Witness Ford notes that the Truncated Z is the only method proposed by the parties to aggregate cell-level statistics. While witness Bell has some concern about Truncated Z concealing discrimination, he notes that "this feature of truncated Z is not a flaw in the procedure, but it can result in unintended consequences if very heterogeneous cells are aggregated."

We agree with the premise that the choice of Truncated Z or Modified Z depends on the level of disaggregation. Fundamentally, the issue is the appropriate level of disaggregation for enforcement measures, with the statistical methodology being a fallout. Based on findings on the appropriate level of disaggregation, the Truncated Z statistic shall be used to evaluate compliance for enforcement measures with retail analogs. For small samples (30 or less), BellSouth witness Mulrow, ALEC Coalition witness Bell, and Z-Tel witness Ford agree that a permutation test should be used to calculate Z-scores for mean measures. Witness Bell explained that permutation analysis is a computer-intensive method that compares the observed results for the ALEC customers with the distribution

of results that would be observed by drawing a random sample from the pool of ALEC and BellSouth customers.

With respect to proportion and rate measures, the testimony evolved over the course of the proceeding, with the ultimate outcome being that there is considerable similarity in the positions being taken by BellSouth witness Mulrow, ALEC Coalition witness Bell, and Z-Tel witness Ford. While witnesses Mulrow, Bell, and Ford acknowledge the "odds" ratio method as being legitimate, witnesses Bell and Ford note that no evidence has been presented regarding the appropriate value for psi, a key parameter of the test. On this basis, we find that the "odds" ratio shall not be considered.

The other method cited for proportion measures and, in some instances rate measures, is the transformed data method, also known as the arcsine square root transformation. BellSouth witness Mulrow, ALEC Coalition witness Bell, and Z-Tel witness Ford all support use of this method to calculate Z scores for proportion measures. Further, witnesses Bell and Ford support use of this method to calculate Z scores for rate measures, while witness Mulrow contends that the square root transformation should be used for rate measures. According to witness Mulrow's testimony, however, he has not verified the appropriateness of using the square root transformation for rate measures, and is relying on a representation made by Dr. Mallows, a former AT&T statistician, who is not a witness in this case. Accordingly, we find that the weight of the evidence supports use of the transformed data method for both proportion and rate measures. In addition, while proportion measures cannot exceed 1, and a rate measure can in theory exceed 1, we find that there will be little practical difference in the range of values for these two types of measures, in the context of a performance assessment plan. For small samples, all witnesses who offered an opinion stated that the hypergeometric test, also known as Fisher's Exact Test, is appropriate for proportion and rate measures.

Based on our findings above, the Truncated Z statistic shall be used to evaluate compliance for enforcement measures with retail analogs. For small samples (30 or less), a permutation test shall be used to calculate Z-scores for mean measures. In addition, the transformed data method, also known as the arcsine square root transformation, shall be used to calculate Z-scores

for proportion and rate measures. For small samples, the hypergeometric test, also known as Fisher's Exact Test, shall be used for proportion and rate measures.

B. Parameter Delta

Witness Stallcup stated that Balancing Type I and Type II errors requires inclusion of a parameter called "delta," which introduces the concept of material disparity. BellSouth witness Mulrow defines delta as "a factor that is used to identify whether a meaningful difference exists between the BellSouth and ALEC performance, in addition to a statistically significant difference." ALEC Coalition witness Bell describes delta as the degree of disparity for which the probabilities of Type I and Type II errors are being balanced. He opines that "this disparity should equal the minimum difference that is judged to be a material obstacle to competition." BellSouth witness Taylor describes delta as a material difference and elaborates that "delta is the number that balances the penalty payment with the gain from discrimination." The parties are in agreement that the choice of a delta value is not really a statistical decision, but rather a decision based on business judgment.

We note that because delta introduces disparity, while at the same time the statistical test should theoretically be one of parity, there is an inherent tension between these two concepts. Z-Tel witness Ford indicates that the larger the value of delta, the further the statistical test deviates from a true test of parity. In exchange for this undesirable result, there is the gain achieved by balancing statistical errors. Witness Ford stresses that the balancing effort should be done in a reasonable fashion in order to minimize the extent to which the statistical test deviates from a true test of parity.

BellSouth is recommending a delta value of 1 for Tier 1 and .5 for Tier 2. To illustrate the practical effect of delta, BellSouth witness Mulrow provides a provisioning example using a measure with a mean of 5 days and a standard deviation of half a day. Using first a delta value of 1, and then a delta value of .5, witness Mulrow indicates that if the ALEC mean exceeds BellSouth's mean by 6 hours and 3 hours, respectively, the differences would be viewed as material. He questions whether such a small difference is really material.

ALEC Coalition witness Bell and Z-Tel witness Ford both question the usefulness of witness Mulrow's example. Witnesses Bell and Ford both believe that witness Mulrow's example is very unrealistic in that the standard deviation for provisioning intervals typically exceeds the mean. Both witnesses cite to Qwest performance results as one basis for their opinion. In addition, both witnesses provide alternative examples, with purportedly more realistic assumptions for the standard deviation. These alternative examples provided by witnesses Bell and Ford result in differences between the ALEC mean and the BellSouth mean of 5 days and 7.5 days, respectively, being judged material. We note that BellSouth could report standard deviations for interim performance measures, but has chosen not to do so. Thus, there is no empirical evidence, specific to BellSouth, regarding the relationship between the mean and standard deviation for different measures.

The ALEC Coalition recommends that we set the delta value no higher than .25. If the delta value is substantially higher than the minimum value needed to reflect materiality, witness Bell indicates that ALECs will face a greater risk of Type II error than BellSouth's risk of Type I error under a parity test. According to witness Bell, this problem is particularly significant for large sample sizes where the balancing critical value is a large negative, which corresponds to a very small probability of Type I error.

Z-Tel witness Ford advocates a delta function, in which delta varies by sample size, as being a reasonable compromise between the positions of BellSouth and the ALEC Coalition. With witness Ford's recommended parameter values, the equation produces a maximum delta value of 1, and a delta value of .051 at an ALEC sample size of 30,000. For a sample size of 175, the delta value is .25. Under the delta function, we observe that the delta value is inversely related to the ALEC sample size.

There is agreement on the ramifications of the choice of the delta value. BellSouth witness Mulrow and ALEC Coalition witness Bell both state that penalties will be paid if the disparity is greater than ½ delta standard deviations. Witness Bell notes, however, that error balancing does not occur at this point.

There is much dispute regarding the relevance of sample size in selecting the delta value. BellSouth witness Mulrow strongly believes that delta should not vary with sample size. In response to questions regarding the Louisiana statisticians' report, which he coauthored, witness Mulrow contends that the statement "sample size matters here too," which appears in the report, merely indicates that sample size affects the balancing critical value. Interestingly, witness Mulrow does reference a portion of the Louisiana statisticians' report which, states the following:

> Using the same value of delta for the overall state testing [Tier 2] does not seem sensible. At the state level we are aggregating over CLECs, so using the same delta as for an individual CLEC would be saying that a "meaningful" degree of disparity is one where the violation is the same for each CLEC. But the detection of disparity for any component CLEC is important, so the relevant "overall" delta should be smaller.

In addition, witness Mulrow is asked about a statement in the report that a "fixed delta might be fine across individual CLECs where currently in Louisiana the CLEC customer bases are not too different." Witness Mulrow maintains that the statement means that a fixed delta might be reasonable if the CLECs serve similar types of customers, and thus have similar types of transactions. He continues to maintain, however, that sample size should not affect the selection of a delta value, and attributes the confusion to a bad job of cutting and pasting.

The ALEC witnesses offer considerable testimony in opposition to the position taken by witness Mulrow. First, Z-Tel witness Ford disputes testimony by witness Mulrow that the decision to use a lower delta value for Tier 2 in Louisiana is related to the masking which can occur in aggregating results across ALECs. Witness Ford contends that the real reason is that sample sizes are inherently larger for Tier 2, and a lower delta reduces the Balancing Critical Value, which protects the integrity of the statistical test of parity.

Witness Ford also believes that there are perverse consequences from balancing with large sample sizes. ALEC Coalition witness Bell also believes that balancing has some

limitations for large samples. Under his proposed delta function, witness Ford maintains that these difficulties are mitigated. In particular, he states:

> The most important aspect of my proposal on the choice of delta is that once the statistical errors get so small that the errors have no real impact on the over or underpayment of penalties, then we should adhere more closely to a strict test of equality because the balancing procedure forces us to deviate from a true test of equality, an undesirable consequence of the approach.

Witness Ford explains that a standard statistical test which does not employ error balancing takes into account the imprecision inherent in an estimate. This imprecision is quite pronounced at small sample sizes, but at large sample sizes, the estimate is much more precise. Failure to consider sample size in setting a delta value results in greater error at large sample sizes than would occur under a standard statistical test.

BellSouth witness Taylor also has concerns which are related to sample size. For small sample sizes, witness Taylor states that balancing results in high Type I error, as well as high Type II error. He believes this is problematic since statisticians typically err on the side of a "not guilty" verdict when samples are small, and therefore, tests are not powerful. With very large samples, very small differences can be detected. On the one hand, the difference may not be material in the sense of having any competitive significance, but the difference may be statistically significant and consistent with discrimination. Witness Taylor indicates that he does not mind using a balancing critical value for any sample size. In fact, he does not have a magic number for sample size, but indicates that the sample size and delta should yield a balancing critical value on the order of 1.5, which equates to a Type I error or significance level of about .05. Under Dr. Mulrow's approach in which sample size is not considered, significance levels could be drastically lower than .05.

With the exception of the appropriate remedy calculation, we find that the appropriate value of delta is the most contentious aspect of the statistical methodology. To make matters more

difficult, there is no established method for setting delta, and the decision is largely one of judgment, albeit there are statistical considerations.

We find that much of the dispute is related to conflicting objectives. BellSouth witness Mulrow states that "those levels of disparity that are lower than the materiality threshold, which is defined by the choice of delta, will not be considered discriminatory." On the other hand, Z-Tel witness Ford believes that delta is a "necessary evil." In exchange for the statistical test deviating from a true test of parity, the ALECs receive the benefit of error probability balancing.

In our opinion, witness Ford advances the correct principle, namely that balancing should be done in a reasonable fashion in order to minimize the deviation from a true test of parity. We recognize that BellSouth witness Mulrow's position that balancing should be done in the same fashion (i.e., fixed delta) across all sample sizes is probably rooted in the idea that since balancing assists ALECs at small sample sizes, it is only fair the balancing disadvantage ALECs at larger sample sizes. We do not find this rationale compelling. We are persuaded by the principle advanced by witness Ford that we should adhere as closely as possible to a strict test of parity, since BellSouth is required to provide non-discriminatory service under the Telecommunications Act of 1996.

We find that Z-Tel witness Ford's delta function and recommended parameter values shall be adopted since this approach will do a better job of achieving our objective than any of the other proposals. Through the delta function, the delta value will be inversely related to the ALEC sample size. This will ensure that balancing will have less practical effect as the sample size increases, minimizing the extent to which the statistical test deviates from a true test of parity. Moreover, witness Ford's delta function covers the range of delta values proposed by the various parties in this proceeding. Finally, and importantly, witness Ford's proposal is inherently applicable to Tier 1 and Tier 2, since delta is a function of sample size.

C. Remedy Calculation

As mentioned previously, this aspect is extremely contentious since BellSouth and the ALEC Coalition have proposed radically different remedy calculations. BellSouth is recommending transaction-based remedies, while the ALEC Coalition is advocating measure-based remedies. Under BellSouth's transaction-based remedy plan, a payment would be made based on some estimate of the number of discriminatory transactions for a measure and the type of measure. Under the ALEC Coalition's measure-based remedy plan, payments would be made based on a finding of discrimination for the measure, which would be independent of the number of transactions and the type of measure. Both plans purport to address the severity and duration of the discrimination, and the ALEC Coalition Plan includes a market penetration adjustment for Tier 2. As will be discussed below, both remedy plans are problematic in certain respects. In addition, no real empirical data has been presented which can serve as a basis for the penalty amounts under either plan. Consequently, most of the criticisms of both plans are theoretical in nature.

Assuming the goal is to ensure that BellSouth has an economic incentive to comply with performance standards, BellSouth witness Taylor believes that the size of the penalty payments should be calibrated to the seriousness of the performance disparities. He goes on to explain that the economic value should be based mostly on business judgment initially and refined based on experience. For those performance disparities that cross the materiality threshold, he believes that the next step should be to determine what portion of the transactions suffer from "statistically significant and material performance disparities." Witness Taylor alleges that BellSouth is the only party that attempts to make such a calculation. Lastly, the number of affected transactions is multiplied by a per-transaction penalty.

Witness Taylor believes that the penalties in the ALEC Coalition plan are "arbitrary, unrelated to performance metrics or transactions, and unrelated to the economic importance of observed performance disparities." While he acknowledges that BellSouth's proposed penalties are in some sense arbitrary, he believes that the BellSouth plan is more rational. In

particular, witness Taylor believes that the BellSouth plan recognizes the type of transaction, the estimated economic seriousness of the violation, and the duration of the violation. In contrast, he believes that the ALEC plan attempts to determine severity based on statistical criteria and does not correlate the size of the penalty with the economic harm. According to witness Taylor, not all functions or performance metrics have the same economic value.

Witness Taylor goes on to discuss the consequences of setting penalties without regard to the economic significance of the disparity. He indicates that a statistical decision rule will not reflect the expected economic gain or loss from the disparity. As a result, one party may attempt to game the system. The witness defines one type of gaming known as moral hazard as follows:

> . . . moral hazard is a form of gaming by which one party to a plan or contract may act in ways -- within the framework of the existing plan -- that allow it to gain an unanticipated competitive or financial advantage at the expense of the other party.

Moral hazard-based behavior could manifest itself in several ways such as rewarding lack of cooperation, maximizing opportunities for unearned income by ALECs, discouraging investment by ALECs, encouraging inefficient entry, and encouraging entrapment. Witness Taylor believes that the "single best protection against gaming is to de-link the size of penalties for specific performance disparities from the statistical methodology used to test for those disparities."

ALEC Coalition witness Bursh criticizes BellSouth's transaction-based remedy payments as minimizing BellSouth's liability when competition is at an "embryonic" level. In addition, Z-Tel witness Ford believes that a transaction-based approach will favor large ALECs.

ALEC witnesses Bell, Bursh, and Ford take great issue with BellSouth's parity gap and affected volume calculations. As will be described below, these witnesses believe BellSouth's approach for determining the number of adversely affected transactions is conceptually flawed. Even BellSouth witnesses Mulrow and Taylor

acknowledge that in estimating the number of discriminatory transactions, BellSouth proposes to estimate the portion of transactions for which disparate service was detected, rather than the number of transactions that did not receive parity service. Witness Taylor mentions that he does not have a better way of doing the calculation and admits that this notion of affected transactions is not a clear concept. Nonetheless, he believes the calculation is "roughly right" in that the resultant penalties should be sufficient to deter discriminatory behavior.

While unsure how to correct the problem, ALEC Coalition witness Bursh believes there is something terribly inappropriate about paying remedies on only a portion of the violations. She cites an example in which there were remedy payments for only 29 of 96 violations.

ALEC Coalition witness Bell states that he does not understand BellSouth's rationale for the affected volume calculation. He goes on to state that under BellSouth's plan, remedies are paid on the number of transactions beyond the point where BellSouth is found out of compliance, rather than beyond parity. Witness Bell believes the proper concept is that once BellSouth is determined to be out of compliance, the question should be how far has BellSouth deviated from parity. To illustrate his concept, witness Bell provides an analogy where a driver is stopped for speeding, traveling 77 miles per hour in a 65 miles per hour zone. While speeders may not be stopped unless they are going at least ten miles an hour over the limit, the fine is predicated on the driver being 12 miles per hour over the limit. He believes that BellSouth's parity gap calculation is analogous to only being judged out of compliance by two miles per hour.

Z-Tel witness Ford finds the parity gap calculation problematic in several respects. First, he provides two examples in which the average time in which BellSouth provides service to the ALEC is the same, but the distribution about the average is quite different. The parity gap is the same for both examples, but in one case 10% of the transactions are actually discriminatory, while in the other case, all of the transactions are discriminatory. Witness Ford belives it is very odd that the parity gap calculation would produce the same result when the form of discrimination is so different. He also notes that

BellSouth's proposal to truncate the parity gap at 100% is further evidence that the parity gap cannot be a measure of transactions. If the parity gap truly measured transactions, the parity gap could not exceed 100%, and there would be no reason for the truncation. Finally, witness Ford states that "(e)xactly what the parity gap does measure is unclear, particularly after the truncation procedures, but it does not appear to be a reliable measure of either transactions or severity." He believes that the parity gap may indicate discrimination or just differences based on sample size. Furthermore, witness Ford believes that the parity gap is "not a reliable or consistent measure of how far the means are apart."

Under the ALEC Coalition's plan, the maximum penalty per measure for Tier 1 is $25,000 for severe or chronic (three consecutive misses) violations, and the minimum penalty is $2,500. Tier 2 penalties are variable multiples of the Tier 1 penalties, which depend on ALEC market penetration. The penalties are not sensitive to the type of measure.

Z-Tel witness Ford supports the ALEC Coalition's proposal for measure-based remedies since he believes that the decision is to discriminate, rather than to discriminate against certain customers. ALEC Coalition witness Bursh believes that the penalty amounts should incent BellSouth to comply. According to witness Bursh, the ALEC Coalition's proposed penalties are designed to provide the appropriate incentive and are not intended to reflect the economic harm to the ALEC, which she believes is nearly impossible to determine.

In addition to his previous commentary on the "arbitrary nature" of the ALEC Coalition's proposed penalties, BellSouth witness Taylor also criticizes the plan on the basis that the statistical certainty of discrimination is not an indicator of severity. He believes that a statistical decision rule can only provide an absolute diagnosis, not a relative one. Stated differently, the statistical decision rule merely indicates that the null hypothesis is true or false. The statistical decision rule can detect material discrimination, but cannot determine the relative severity of the failure.

Witness Taylor explains that "a z-score that is twice as distant from a critical value than another could easily be for

reasons other than simply that one of the performance means is twice as large as the other." According to witness Taylor, z-scores are influenced by "the mean performance when BellSouth serves itself, the mean performance when BellSouth serves the ALEC, the standard deviations for both, and the number of measurements made in each case."

By using the same method to detect discrimination and measure its severity, witness Taylor believes that the ALEC Coalition's Plan confuses the degree of certainty with the degree of severity. Even ALEC Coalition witness Bursh acknowledges that the penalties escalate as the statistical certainty of discrimination increases. We agree with BellSouth's witness Taylor's assessment that the statistical decision rule is not helpful in assessing severity.

Unfortunately, both the BellSouth remedy plan and the ALEC Coalition remedy plan appear to do a poor job of estimating the extent of any discrimination. As discussed above, the BellSouth plan is predicated on parity gap and affected volume calculations that are very questionable, and the ALEC Coalition plan confuses statistical certainty with severity. Witness Stallcup does note that apart from the level of disaggregation affecting the statistical evaluation, the best parts of both plans could be combined into some sort of hybrid remedy plan. ALEC Coalition witness Bell also observes that a different remedy plan, other than the one proposed by BellSouth, could be used with the truncated Z.

Because the evidence demonstrates that there are fundamental flaws in both the BellSouth and ALEC Coalition remedy plans, we have no choice but to require a remedy plan which incorporates the better features of the two. First, we find that the remedy plan must, at least initially, be measure-based given what we believe to be serious issues with BellSouth's parity gap and affected volume calculations. Over time, it may be possible to evolve to a transaction-based system, with a minimum payment, an idea mentioned by Z-Tel witness Ford. If the issues with BellSouth's parity gap and affected volume calculations can be solved through the periodic review process, we believe that transaction-based remedies, with a minimum payment provision, would be preferable in concept. For now, however, we see no

choice but to require that a measure-based remedy plan be adopted.

We note that BellSouth's recommended remedy payment per affected item varies depending on the measure, while the ALEC Coalition's recommended remedy payment per failed measure does not vary according to the type of measure. In concert with BellSouth witness Taylor's testimony, we find that economic importance is a relevant consideration in setting remedy payments. By the same token, we acknowledge ALEC Coalition witness Bursh's testimony, that the economic cost to ALECs is almost impossible to pinpoint. In addition, we find that certain measures are intrinsically more important in that success or failure in meeting the standard more directly affects end use customers.

Based on the above considerations, we find that the remedy payments shall vary by measure. Unfortunately, no empirical evidence was offered by any party to this proceeding, which can be used to set remedy payments. As a result, the relative relationships between the various BellSouth proposed remedy payments provide the only quantitative basis for differentiating remedy payments by measure.

BellSouth and the ALEC Coalition both address chronic failures, but in slightly different ways. Under the ALEC Coalition Plan, a chronic failure is defined as three consecutive monthly misses and calls for a $25,000 payment under Tier 1. BellSouth proposes a sliding scale of remedy payments for Tier 1, in which the penalty increases for successive months of non-compliant performance.

BellSouth proposes separate schedules of remedies for Tier 1 and Tier 2. Also, under BellSouth's proposal, Tier 2 penalties are assessed after three consecutive months of violations. In contrast, the ALEC Coalition recommends that Tier 2 remedies be a multiple of "n" greater than the Tier 1 remedies. The value for "n" is a function of the ALEC market penetration levels and varies from 1 to 10.

Given our requirement to vary remedy payments by measure, and in view of the fact that BellSouth's recommended remedies, per affected item, vary by type of measure, tier, and duration,

we find that these relationships could be used to deaverage the ALEC Coalition's recommended $2,500 minimum payment per failed measure.

In general, the easiest way to implement this concept would be to apply a multiplier to BellSouth's remedy tables for Tier 1 and Tier 2 to convert to measure-based penalties. A problem will arise, however, for certain measures where the volumes are expressed in very different units, as compared to other measures. For most measures, the volumes are expressed in terms of end user orders. This is true for pre-ordering, ordering, provisioning, maintenance and repair, and LNP. This is not the case for billing, change management, interconnection trunks, and collocation.

Based on the above considerations, BellSouth shall develop a remedy plan which includes certain features. Remedies shall be measure-based, rather than transaction-based, and shall vary by type of measure and duration for Tier 1, and type of measure for Tier 2. The relative relationships between the various measure-based remedy payments shall be consistent with the relative relationships between the various BellSouth proposed, transaction-based remedy payments. Tier 1 remedies shall be set such that the average Month 1 remedy approximates the $2,500 minimum payment recommended by the ALEC Coalition. Tier 2 remedies shall be applicable after three consecutive months of violations, as proposed by BellSouth.

We are requiring approximately 825 levels of disaggregation for Tier 1 compliance reporting and penalties. Further, at the time of the hearing in this docket, 92 ALECs had access to Florida PMAP data. Assuming an average remedy payment of $2,500 for Month 1, various scenarios of total monthly payments by BellSouth under Tier 1 can be developed. One awkward aspect of developing scenarios, however, is that the typical ALEC will have transactions in only some of the 825 levels. If the typical ALEC has transactions in only 100 levels, which we believe is a high-end estimate, and there is a 10% failure rate, BellSouth's total monthly payment for Tier 1 would be $2,300,000 ((92 ALECs)(100 levels)(10%)($2,500 average)).

D. Benchmark Table for Small Sample Sizes

With small samples, the parties agree that some consideration must be given to random variation which may make it difficult for BellSouth to meet a benchmark which is expressed as a certain percentage of transactions being completed in a specified time. BellSouth witnesses Coon and Mulrow advocate a statistical approach based on a 95% confidence interval. ALEC Coalition witness Bursh advocates a non-statistical approach, wherein the allowable number of missed transactions is rounded up to the next whole number. For example, in the case of four transactions, an adjustment would be made to allow BellSouth to miss one transaction and still be considered in compliance with a 95% benchmark. Witness Bursh believes that this approach is appropriate because some mitigation has been provided by defining the benchmark at 95%, rather than 100%. Z-Tel witness Ford also believes that there should not be any statistical adjustments to the benchmarks in the case of small sample sizes. He believes that the rounding up approach sponsored by witness Bursh is reasonable.

Since the benchmarks are set in a way that does not require perfection, we find that the rounding up approach advocated by ALEC Coalition witness Bursh could be used. We note that such an approach would be simpler. Nonetheless, we find that BellSouth's recommended approach is more defensible since it incorporates random variation in a statistically sound manner. To illustrate the difference in the two approaches, consider the case where there are 20 transactions for a particular measure, and the benchmark is expressed as 95% of the transactions being completed in 24 hours. If 18 of the 20 transactions (90%) are completed in 24 hours, this would be considered non-compliant performance under the ALEC Coalition Plan, and compliant performance under BellSouth's Plan. We believe that BellSouth's approach takes into consideration that its typical performance can meet the 95% standard, yet be higher or lower for a small sample because of random variation. On this basis, we find that BellSouth's recommended benchmark table shall be adopted for small samples.

E. Floor on the Balancing Critical Value

ALEC Coalition witness Bell and Z-Tel witness Ford both believe that there should be a floor on the balancing critical

value in certain situations. Witness Bell supports use of a floor if the delta value is greater than .25 and also sees merit in using either a floor for large sample sizes or the delta function. Similarly, witness Ford believes that a floor is needed, or the delta value should be a function of sample size.

In contrast, BellSouth witness Mulrow does not believe a floor is appropriate since this would artificially and arbitrarily reduce the materiality level. He explains that when sample sizes are small, balancing results in significance levels that are much larger than conventionally used, which gives the benefit of the doubt to the ALEC. When sample sizes are large, the reverse is true, and the data should show a material difference, not simply a conventionally significant difference.

In view of our decision to adopt Z-Tel witness Ford's delta function, there is no need to place a floor on the balancing critical value. Indeed, witness Ford acknowledges that either a floor is needed or the delta function should be used. Therefore, based on our decision above, there shall not be a floor on the balancing critical value.

XIX. DUE DATE AND METHOD OF PAYMENTS FOR TIER 1 AND TIER 2 NONCOMPLIANCE

As a backdrop, we note that the parties presented relatively little testimony on this issue.

Witness Stallcup and BellSouth witness Coon provide similar proposals. According to witness Stallcup, payment should be made by the 30th day following the due date of the performance measurement report for the month in which the obligation arose. Witness Coon believes that payment should be made by check, by the end of the second month following the month for which disparate treatment was detected. The essential difference in the two proposals is that witness Stallcup believes that performance measurement reports should be due by the 20th calendar day of the month, whereas witness Coon believes that the reports should be due by the 30th calendar day of the month, for the preceding month. Both witnesses advocate roughly a month between the due date for the reports and the due date for payment of any obligations arising from the reports. Finally, ALEC Coalition witness Bursh believes that payments for Tier 1 and

Tier 2 noncompliance should be made by the 15th business day following the due date for the reports.

Based on the limited testimony, we find that there is more sentiment towards having a month or 30 days between the due date for the reports and the due date for payment of any obligations arising from the reports. Given that the number of days in a month can vary between 28 and 31, we prefer that the interval be expressed as 30 days. Finally, we note that the parties agree on making payments by check.

Therefore, we find that BellSouth shall make payments for Tier 1 and Tier 2 noncompliance by check, by the 30th day following the due date of the performance measurement report, for the month in which the obligation arose.

XX. INTEREST ON DELINQUENT TIER 1 PAYMENTS

We find it appropriate to approve the following stipulated position, which was agreed to by BellSouth, AT&T, e.spire, FCTA, Worldcom, KMC, Covad, Mpower, Z-tel, Time Warner and IDS, and filed in this docket as document number 09141-01.

> BellSouth shall pay the ALEC interest at a rate of six percent simple interest (at a rate of six percent simple interest per annum) for each day after the due date that BellSouth fails to pay the ALEC.

XXI. FINES FOR DELINQUENT TIER 2 PAYMENTS

In this Section, we address whether BellSouth should be held liable for failure to make payments by the due date under the Tier 2 enforcement mechanism.

Arguments

In its brief, BellSouth argues that the ALECs' position is unnecessarily complex as well as arbitrary. BellSouth further points out that, in Florida, BellSouth is no longer subject to rate of return regulation, but rather to the form of alternative regulation described in Section 364, Florida Statutes. BellSouth contends that the ALEC proposal not only contains an overly

complex calculation, but also that it is based on an anachronistic view of the status of regulation in Florida.

BellSouth witness Coon proposes that "BellSouth make a voluntary payment to the Commission of $1,000 per day for each day after the due date that BellSouth fails to pay under the Tier 2 Enforcement Mechanism." With the exception of BellSouth's payment being voluntary as opposed to an involuntary penalty or a fine, witness Stallcup agrees with BellSouth's proposal. Both witnesses agree that $1,000 per day is appropriate and should be deposited into the State General Revenue Fund.

Witness Bursh states, "[i]f the ILEC fails to remit a consequence payment . . . then it should be liable for accrued interest for every day the payment is late." She further states that the interest should be calculated at "[a] per diem interest rate that is equivalent to the ILEC's rate of return for its regulated services for the most recent reporting year." However, in its brief, the ALEC Coalition states, "[i]nterest should be calculated in the same manner as the late payment for Tier 1 measures." As stated above, the parties to this docket stipulated that BellSouth would pay the ALECs interest at a rate of six percent simple interest per annum for each day after the due date for the Tier 1 enforcement mechanism.

DECISION

Based upon the evidence presented, we concur with BellSouth's position. It is unclear as to which method of payment the ALECs prefer: a per diem interest rate equivalent to BellSouth's rate of return or the stipulated six percent simple interest.

As asserted by BellSouth in its brief, BellSouth is no longer subject to rate of return regulation in Florida. Hence, it is not possible to set an interest rate equivalent to BellSouth's rate of return.

We also find the calculation of a six percent simple interest rate would be unnecessarily complex. The ALECs would not benefit from customizing each payment amount since the payments under the Tier 2 enforcement mechanism would be made to us for deposit in the State's General Revenue Fund. As observed